



ISSN 2179-4529 – ANAIS DO 4º SIMPÓSIO DE CIBERJORNALISMO

O uso do Data Mining para descoberta de enquadramentos durante a eclosão das manifestações nos sites G1 e Estadão

Pedro Alexandre Cabral

Thamirys Viana

Resumo

As notícias veiculadas pelos websites jornalísticos são abordadas através de diversos enquadramentos e temas, disponibilizando um volume cada vez maior de informações aos leitores a respeito de determinado acontecimento. Nesse contexto, o presente trabalho tem como objetivo demonstrar como os processos de *Data Mining* (TURBAN, 2005) podem ajudar na descoberta dos enquadramentos (TUCHMAN, 1978), (SÁDABA 2007) noticiosos nos sites de referência G1 e Estadão, durante a eclosão das manifestações populares, no período de 20 a 30 de julho deste ano. A partir da identificação e quantificação das palavras-chave, pretende-se descobrir quais termos foram mais proeminentes e como o acontecimento foi modelado ao longo da cobertura jornalística. A partir dessas informações iniciais, vislumbra-se montar um quadro comparativo entre os dois portais estudados, a fim de descobrir se na construção da notícia existe alguma similaridade em relação aos enquadramentos utilizados. Para detectar a formação de padrões nos conteúdos veiculados utilizaremos a ferramenta Solr e um algoritmo de mineração desenvolvido, especificamente, para este fim.

Palavras-chave: Enquadramento. Data Mining. Manifestações. G1. Estadão.

Introdução

Tendo em vista o imenso volume de informações disponíveis no âmbito da Web e a dificuldade cada vez mais premente em se analisar mananciais de dados, nasce a necessidade de mecanismos que possibilitem a técnica conhecida como mineração de dados. Neste contexto, o presente trabalho busca analisar o enquadramento utilizado pelos



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





portais G1 e Estadão durante a eclosão das manifestações populares que tomaram conta do país no mês de junho deste ano, através de técnicas de *Data Mining*.

Os objetivos secundários visam evidenciar as temáticas abordadas nos portais analisados, bem como comparar os padrões de enquadramento sob a perspectiva do local em que o fato ocorreu, o caráter da notícia, o tempo em que foi noticiada, bem como as temáticas mais recorrentes. Para tanto, foram analisadas 15.116 peças noticiosas, postadas no período de 20 a 30 de junho do corrente ano, contidas em ambos os portais. É válido ressaltar que esse período leva em conta a ocorrência da maior incidência de atos de protesto nas ruas de várias cidades brasileiras.

Como hipóteses para as questões supracitadas temos que o enquadramento relacionado ao quesito local se dará em maior intensidade com relação à região Sudeste; o momento em que a notícia será enquadrada, em sua maioria, estará no tempo presente; além do enquadramento utilizado na maior parte das vezes ter caráter negativo, visto que acreditamos na existência de uma alta frequência do termo vandalismo e derivados dele.

Para conceituar enquadramento entraremos aporte em SÁDABA (2007), TCHUMAN (1978), SILVEIRINHA (2013), enquanto as discussões e conceitos sobre *Data Mining* terão como base GOLDSCHMIDT & PASSOS (2005), PINHEIRO (2009), TURBAN (2005) e SOARES (2007).

O presente artigo justifica-se devido à noção de enquadramento jornalístico ser um requisito que demonstra os contornos percorridos pelos meios de comunicação quando da construção social da realidade. Além da premente necessidade de extrair informações até então desconhecidas, utilizando como base técnicas de *Data Mining* e assim entender como se dá a relação dos métodos adotados por esta perspectiva quando do trato de informações jornalísticas.



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





Teoria do Framing: evolução, conceitos e perspectivas

Grande parcela da população mundial só toma conhecimento dos fatos importantes do mundo ou mesmo aqueles que ocorrem nas proximidades em que vive, pois existiu ali um comunicador, que registrou o acontecimento e o relatou através de um meio de comunicação. Tais fatos vão ser narrados, segundo a ótica do jornalista contemporâneo, o qual foi orientado a perseguir parâmetros como a objetividade, clareza, concisão, dentre outros adjetivos que fazem com que a notícia se transforme numa possível “transcrição do real”, possibilitando que o leitor tenha condições de construir sua visão de mundo a respeito de determinado acontecimento.

Por notícia, Erbolato (1991) entende que seja a matéria-prima do Jornalismo, visto que “somente depois de conhecidas ou divulgadas é que os assuntos aos quais se referem podem ser comentados, interpretados e pesquisados” (ERBOLATO, 1991, p. 49). Apesar de ser um dos formatos jornalísticos mais difundidos na sociedade, até hoje não existe um consenso sobre a definição do que é notícia. Estudiosos da questão apontam alguns valores que devem vir incutidos neste formato, sendo intitulados de valores-notícia ou ainda critérios de noticiabilidade.

Nesse contexto, para entender o surgimento e aplicação do Framing é necessário refazer um percurso que inicia há muitos anos no âmbito de outras disciplinas. Sádaba (2007) relata que a teoria do enquadramento (Framing) surgiu a partir das preocupações de psicólogos e sociólogos com os modos de conhecimento. Contudo, com o passar do tempo, esta teoria foi alcançando domínios científicos diversos, em especial os estudos voltados para os meios de comunicação. Parte-se da concepção sociológica de “definição da situação” (Idem, p. 24), em que as ocorrências tidas como reais vão ser verdadeiras também em suas consequências. Em outras palavras, a realidade que for interpretada (aqui se faz referência aos meios de comunicação), vai se tornar “la realidad social por excelência” (Ibidem, p. 25). Tomando de empréstimo à Psicologia, outro conceito também se tornou importante para compreender futuramente o que viria a ser a teoria do enquadramento: a



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





ideia de “marco” (BATESON, 1972). Ou seja, compreender o motivo pelos quais as pessoas priorizam determinados aspectos da realidade em detrimento de outros.

Assim, tomando como referência as ideias de Entman (1993 *apud* GONÇALVES, 2005, p. 162) o termo enquadrar seria, essencialmente, uma forma de seleção e enaltecimento de determinada proeminência presente no acontecimento central. O framing se configura como uma forma de selecionar determinados aspectos socialmente percebidos, tonando-os mais vistosos ao olho do leitor, partindo da definição de um problema, sua interpretação, avaliação ou ainda a recomendação para determinado fato em questão. A saber,

Os enquadramentos são princípios de seleção, ênfase, e de apresentação [...] compostos por pequenas teorias tácitas sobre o que existe, o que ocorre e o que é relevante. (...) Os enquadramentos mediáticos são padrões persistentes de cognição, interpretação e apresentação, seleção, ênfase e exclusão, através dos quais os manipuladores de símbolos organizam rotineiramente o discurso, seja verbal ou visual (GITLIN *apud* GONÇALVES, 2005, p. 160).

Partindo deste conceito, qualquer ação pode ser entendida por meio dos frames, que apresentam, em sua maioria, uma predominância da sociedade sobre o indivíduo, devido ser entendido como um processo de construção social da realidade. Em outras palavras, o framing busca abarcar, justamente, as respostas para os modos como os homens conhecem o que ocorre a sua volta, dando-lhe um significado. Entretanto, mesmo considerando que os meios de comunicação ocupam hoje uma posição dominante quando do fornecimento de conhecimento à sociedade e organização da realidade social, alguns autores tecem críticas em relação aos enquadramentos utilizados pela mídia, visto que eles não são o único meio que o cidadão possui para informar-se (SÁDABA, 2007, p. 43).

Os princípios que regem a teoria do enquadramento, por sua vez, estarão orientados por uma série de convenções, que vão desde percepções pessoais a cerca de um tema, bem como fatores como proximidade dos acontecimentos, meios disponíveis para a cobertura do



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





fato ou ainda os constrangimentos organizacionais, que são expressos, mais claramente, pela linha editorial que o meio de comunicação põe em prática. Os relatos têm que encontrar apelo junto aos públicos, sendo que o esforço pessoal do jornalista também será um fato digno de destaque. Para Tuchman (1978, p.4),

[...] los elementos que determinan lós frames de lós medios de comunicación sons su organización y lãs actitudes profesionales de lós periodistas que trabajan em ellos. [...] los marcos serían lãs normas transmitidas y compartidas por lós miembros de uma redacción a través de las cuales se mira la realidad sobre la que se informa. La noticia se constituiria como um proceso de socialización, ya que daria forma a lós sucesos de modo que éstos pudieran ser compartidos socialmente y se definiria con “um producto de lós informadores que actúan dentro de procesos institucionales y de conformidad com prácticas institucionales.

A análise dos enquadramentos noticiosos na prática se dará através de questões como o uso da linguagem, visto que a escolha de determinados vocábulos implicam um juízo de valor em relação aos acontecimentos (SÁDABA, 2007, p. 64). No presente trabalho, termos como manifestantes versus vândalos, por exemplo, vão designar circunstâncias diferentes, denotando uma impressão positiva para a primeira, enquanto o uso da segunda transparecerá uma ideia negativa. É válido ressaltar que a autora argumenta que o uso de determinadas expressões também vão depender dos contextos culturais de cada local. Mais uma vez é possível perceber que a teoria dos enquadramentos rompe com a lógica da objetividade, já que a escolha dos vocábulos a serem utilizados fica a cargo do jornalista. Tuchman (1978), enumera outros elementos que interferem no enfoque dado à notícia, a saber: lugar de cobertura, organização em seções, as pessoas que selecionam a notícia, momento em que a notícia ocorre e quando esta é relatada, além de suas tipificações. Outros critérios para análise dos enquadramentos poderão ser utilizadas durante as análises deste artigo.

Sádaba (2007) reúne ao final da compilação Framing: El encuadre de las noticias, quatro proposições para entender os meios de comunicação, por meio da perspectiva



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





proposta pela Teoria do Framing. A primeira delas centra-se no jornalista, enquanto indivíduo que interfere diretamente na produção do conteúdo noticioso. Nesse contexto, fatores como a etnia, o sexo, a educação recebida, bem como as atitudes profissionais e pessoais podem acarretar diferentes perspectivas para a peça noticiosa. A segunda proposição centra-se nas rotinas profissionais, que englobam os modos como os profissionais trabalham nos meios de comunicação. O uso da pirâmide invertida, por exemplo, se configura como uma tendência a ser seguida por determinado meio, enquanto outro pode optar por novas formas de produção da notícia, respeitando as necessidades impostas pelo tipo de veículo no qual está inserido. A organização do meio de comunicação também figura como uma vertente de estudo, tendo em vista as estruturas hierárquicas dentro de uma redação, por exemplo. E por último atenta-se ao conteúdo das notícias, a partir das quais poderão ser analisadas os enquadramentos dados pelos profissionais da Comunicação.

A saber, o presente trabalho segue uma linha de pesquisa baseada na análise do conteúdo das notícias publicadas, a partir das quais vão ser observados enquadramentos, segundo as seguintes variáveis: local em que o fato ocorreu, o caráter da notícia, o tempo em que foi noticiada, bem como as temáticas mais recorrentes.

Data Mining: aplicações e perspectivas

Os constantes avanços na área da Tecnologia da Informação (TI) têm viabilizado o armazenamento de grandes e múltiplas bases de dados (GOLDSCHMIDT & PASSOS, 2005). Nesse contexto, analisar Base de Dados (BDs) não estruturadas como textos sempre representou um entrave aos profissionais, em especial os da Comunicação, devido a grande dificuldade de sistematizar o significado do que lá está escrito. Signos gramaticais, como acentos e abreviações, por exemplo, promovem significações adversas, inserindo uma contextualização diferente. (PINHEIRO, 2009).



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





Deste modo, surge a Mineração de Dados ou *Data Mining* (DM), que configura-se como ativo capital para descoberta de novas informações em textos. Turban (2005) aponta que a técnica dá subsídios para que o usuário possa encontrar o conteúdo “oculto” dos documentos, incluindo relacionamentos úteis e adicionais. Weiss (2007, p. 49) esclarece o conceito de *data mining* como sendo:

Busca de informação valiosa em grandes volumes de dados. Data Mining é o esforço desenvolvido por homens e máquinas. Os homens desenham os bancos de dados, descrevem os problemas e setam os objetivos. As máquinas mineram os dados, em busca de padrões que atendam a estes objetivos.

A partir desta perspectiva, Ebecken et al. (2003) enumera dois tipos de abordagem que podem ser utilizadas, a saber: análise semântica e análise estatística. A análise semântica, por sua vez, é pautada na abordagem natural de leitura, ou seja, como os seres humanos interpretam um texto, fazendo uso do significado das palavras, características morfológicas, sintática e do contexto em geral, mantendo o foco na funcionalidade dos termos do texto. (CARRILHO 2007).

A análise estatística, por sua vez, pauta-se na frequência em que cada termo aparece no texto, sendo que “informações sobre contextualização, precedência ou sucessão de outros termos não são consideradas” (SOARES 2007 p 43). A principal vantagem desta análise é que pode ser utilizada em qualquer idioma. A contextualização do termo, bem como em que parágrafo está inserindo, que termos o antecedem ou que estão diretamente relacionados são irrelevantes para este tipo de abordagem.

A predominância da informática nas rotinas produtivas do Jornalista tem levado ao uso da expressão Jornalismo Digital em Bases de Dados (JDBD), visto que o computador deixou de ser um mero assistente na produção das peças noticiosas, para se tornar um elemento estruturante da própria notícia (TRÄSEL, 2009). As tendências atuais apontam para o uso do *data mining* no jornalismo em Base de Dados (LIMA JR., 2004; FIDALGO, 2007). Essa técnica baseia-se no tratamento matemático de grandes volumes de dados para



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





a identificação de padrões inesperados. Diferencia-se da pesquisa em bancos de dados por não buscar uma resposta para uma pergunta específica, mas sim respostas para perguntas que ainda não foram feitas. Assim, o DM pode dar grandes contribuições ao jornalismo, especialmente na formulação de pautas inovadoras e na descoberta de fatos sociais, políticos e econômicos desconhecidos – padrões encobertos por números.

Para Barbosa *et.al* (2007) é importante ressaltar que apesar do Data Mining ser amplamente discutido no campo da ciência da computação, o esforço de relacioná-lo em aplicações no Jornalismo é recente. Portanto, naturalmente, enfrenta dificuldades de compatibilização. Entretanto, o processo de mineração identifica por meio de tarefas e técnicas as perguntas e respostas na base de dados. Em síntese, é possível não só relacionar eventos com base no histórico, mas a partir daí, atuar de modo preditivo. Outra aplicação também relevante para a área é a Mineração de Textos (MT). De acordo com Tan (1999), esta nova área é definida como o processo de extração de padrões ou conhecimentos, interessantes e não-triviais, a partir de um conjunto de documentos textuais.

A conjunção destes fatores técnicos e socioeconômicos provocam reconfigurações profundas nas rotinas produtivas e na profissão do jornalista. O conhecimento das técnicas de jornalismo em *data mining* se torna cada vez mais essencial e a habilidade para o tratamento de dados pode ser uma vantagem na busca de novos públicos, uma vez que confere credibilidade e permite a descoberta de “furos” e pautas inovadoras. No jornalismo, é grande o volume de dados guardados em arquivos históricos e, na internet, temos acesso a banco de dados dos mais variados temas. Sérgio Navega (2002) relata que, talvez a forma mais nobre de se utilizar vastos repositórios, seja tentar descobrir se há algum conhecimento escondido neles. Segundo com Fidalgo (2007, p.134), no que tange a aplicabilidade do DM no jornalismo, encontramos três vantagens principais:

A primeira é não exigir tanto tempo de trabalho de jornalistas como uma análise interativa (baseada em consultas individuais escolhidas caso a caso) de um grande volume de dados pode implicar. A segunda é um ganho de capacidade de extração de conhecimento, devido a ser possível



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





testar um número muito mais vasto de hipotéticas relações interessantes ao nível dos dados do que seria possível se essas hipóteses tivessem de ser diretamente sugeridas ao sistema de pesquisa (e os seus resultados tivessem de ser diretamente avaliados) por especialistas humanos. A terceira vantagem, mais subtil, é que ao testarem automaticamente conjuntos exaustivos de possíveis padrões, os sistemas de KDD (*knowledge-discovery in databases*) tem mais probabilidades de testar e encontrar conhecimento inesperado (e inovador) do que se pode esperar de consultas interativas sugeridas com base no conhecimento sobre o domínio previamente existente.

Metodologia

O presente trabalho sintetiza os preceitos de uma pesquisa quantitativa, visto que pretende-se numerar, medir unidades e estabelecer relações entre os enquadramentos noticiosos. O objetivo deste tipo de trabalho é buscar padrões através de ferramentas de mineração de texto, presente no conteúdo dos portais de referência G1 (www.g1.com) e Estadão (www.estadao.com.br), no período de 20/06/2013 a 30/06/2013, período em que eclodirão as manifestações populares em todo o país, que dentre as motivações constavam a abolição a corrupção, redução do preço da passagem de ônibus, bem como a busca de outros direitos, a exemplo do reconhecimento da união homossexual.

Para atender aos objetivos propostos nesta pesquisa, utilizaremos o processo estabelecido por Aranha C. N (2007): coleta de dados, pré-processamento, indexação, mineração e análise. A primeira fase será a coleta de notícias, cujo objetivo é montagem de um banco de dados, elemento básico de qualquer processo de Data Mining. O referido banco foi alimentado através de um algoritmo que acessou os dois portais de estudados e em seguida salvou todas as noticias coletas na ferramenta Solr.



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com



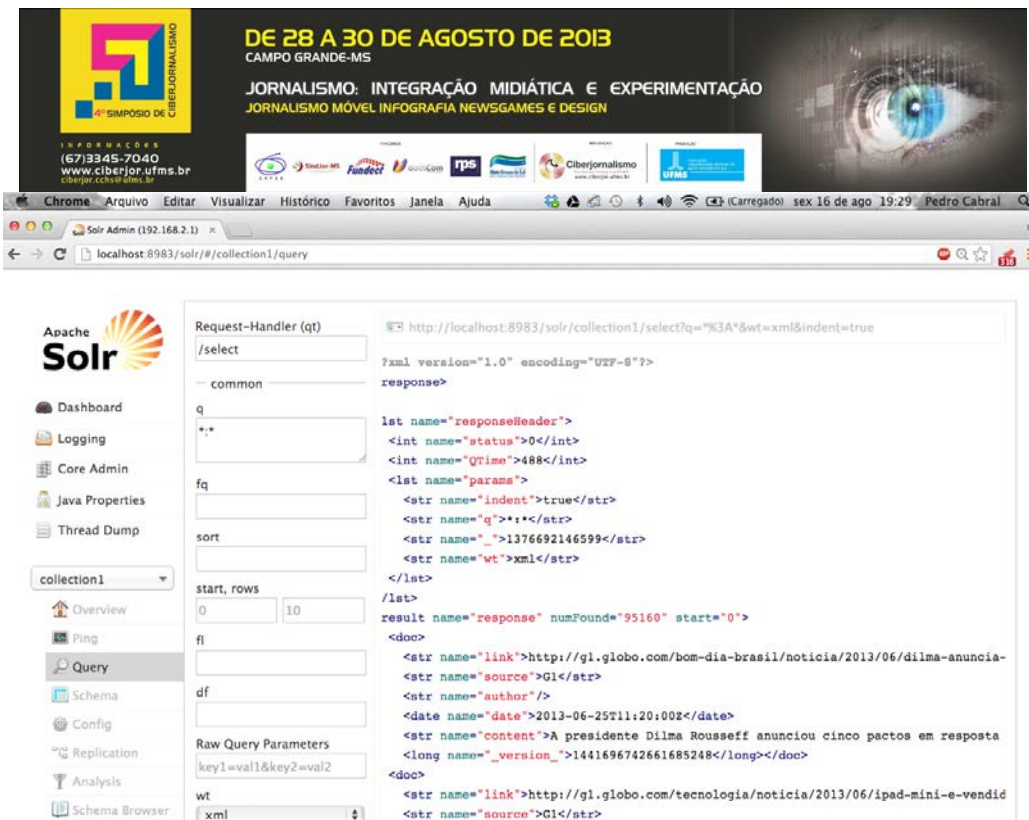


Figura 1 – Ferramenta Solr

A partir dos dados coletados foi estruturado um novo banco de dados para realizar o pré-processamento, as notícias foram armazenadas no Solr de acordo com os seguintes atributos: link (url da notícia), source (portal da notícia, G1 ou Estadão), author (repórter que escreveu a notícia), data (data da notícia), title (título da notícia), e content (conteúdo da notícia). Uma vez realizada a coleta de dados, o próximo passo foi a preparação dos textos para que os mesmos possam ser manipulados pelos algoritmos de Mineração de Textos. Nesta etapa, as notícias armazenadas no banco de dados foram submetidas a inúmeras operações capazes de obter uma forma de representa-los estruturadamente. Após o pre-processamento iniciou-se a fase da indexação, que é o processo responsável pela criação de estruturas auxiliares denominadas índices e que garantem a rapidez e agilidade na recuperação das notícias, esse processo foi feito de forma automática pela ferramenta Solr.

A quarta fase é a etapa de mineração, onde ocorre a busca efetiva por conhecimentos novos e úteis a partir dos dados. Compreende a aplicação de algoritmos sobre as notícias



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
 Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
 CEP 79070-900 * Campo Grande (MS) *
 www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





coletadas de forma a abstrair o conhecimento implícito e buscar padrões. Finalizado o processo de mineração de textos, tivemos a etapa de análise. A etapa de Análise, algumas vezes chamada de Pós- Processamento, abrange o tratamento do conhecimento obtido na etapa de Mineração, através da análise, visualização e interpretação dos dados. Para auxiliar o processo de análise foi exportado um arquivo do tipo CSV (Comma-separated values) e submetido a ferramenta pivot table do Microsoft Excel for Mac 2011.

Análise

Durante o período de 20/06/2013 a 30/06/2013 foram contabilizadas 15.116 notícias. Ao refinar a busca pela ocorrência da palavra “protesto”, a quantidade reduziu para 5.765. Ou seja, 38% das notícias veiculadas no período traziam em seu bojo a pauta “protesto”, sendo que destas, 82% foram veiculados no G1 e outros 18% no Estadão. A disparidade entre a quantidade de peças noticiosas entre ambos os portais justifica-se pelo fato do veículo G1 possuir uma “central de reportagem” mais capilarizada. Em outras palavras, várias seções estão interligadas a este portal, onde são pontuadas notícias de cunho estadual, regional, nacional e mundial, além dos cadernos que são disponibilizados dentro da programação da Rede Globo.

Na tentativa de descobrir enquadramento por local, após a coleta da notícias foi iniciado o processo de data mining buscando quantificar a relação nome do estado e o termo “protesto”, como resultado obtivemos para a região Sudeste, 3.174 notícias. O estado de São Paulo foi responsável por 32% das ocorrências encontradas, seguido por Minas Gerais (26%), Espírito Santo (22%) e em ultimo o estado do Rio de Janeiro com 19% das ocorrências.

Já na região Nordeste, o mesmo parâmetro utilizado na busca, nome do estado e a palavra “protesto”, a quantidade de notícias encontradas foi bem menor 691 notícias, sendo que o estado do Rio Grande do Norte apresentou maior quantidade 46% (ver tabela3).



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





Dentro dessa perspectiva, podemos inferir que o enquadramento com a temática protesto nos veículos analisados, G1 e Estadão, deu-se de maneira mais profunda na região sudeste.

Estado	Palavra Protesto	%Palavra Protesto
Espírito Santo	713	22%
Minas Gerais	824	26%
Rio de Janeiro	608	19%
São Paulo	1.029	32%
Total	3.174	

Table 2 – Ocorrência da palavra protesto por cidade – Região Sudeste

Estado	Palavra Protesto	%Palavra Protesto
Alagoas	29	4%
Bahia	79	11%
Ceará	55	8%
Maranhão	53	8%
Paraíba	55	8%
Pernambuco	60	9%
Piauí	13	2%
Rio Grande do Norte	320	46%
Sergipe	27	4%
Total	691	

Table 3 – Ocorrência da palavra protesto por cidade – Região Nordeste

Para analisar o momento de notícia, ou seja, o período em que o fato ocorre e a velocidade com que os veículos postam as notícias nos seus websites, foram escolhidas algumas palavras chaves que denotam temporalidade (agora, hoje e ontem). Em seguida foi mapeada a frequência com que cada palavra aparece nos veículos estudados, conforme gráfico 1.

Gráfico 1 - Momento da notícia nos portais

Com base nesses dados, o portal Estadão demonstrou um apreço pela cobertura instantânea, face os 37% da frequência do termo “agora” em suas notícias. Já o portal G1 apresentou, em sua cobertura noticiosa, uma frequência bem menor da referida palavra,



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
 Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
 CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





contabilizando cerca de 50% a menos que o Estadão. Dentro conjuntura, pode-se perceber que o portal Estadão utiliza em sua cobertura alicerces que coadunam com a celeridade imposta pela internet.

A palavra “hoje” também foi muito presente na cobertura do Estadão e mais uma vez o G1 apresentou uma frequência na utilização do termo cerca 50% menor. Entretanto, o portal G1 apresentou uma baixa frequência da palavra ontem, 4%. Isso remete a ideia que embora sua cobertura não seja pautada no imediatismo da notícia, como detectou-se no portal Estadão, ele apresenta uma atualização razoável e não se utiliza de termos que apresentem a ideia de passado.

No que se refere ao enquadramento das temáticas reivindicadas pela população foram segregados os termos com maior frequência nas notícias, em seguida foi realizado um comparativo entre os dois portais a fim de identificar se existia uma similaridade entre os temas pleiteados pela sociedade.

Gráfico 2 - Temáticas Reivindicadas

Percebe-se que, com exceção dos termos transporte e política, os demais apresentaram uma similaridade na frequência dentro das notícias. Nesta conjuntura, o termo com menor frequência foi a PEC 37 (Projeto de Emenda Constitucional que limitaria o poder de investigação criminal a polícias federais e civis), para os dois portais. Embora o item “transporte” apresente uma disparidade acentuada na cobertura das notícias entre ambos os portais analisados, o termo foi o mais recorrente em ambos.



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





No que tange ao caráter da notícia, que neste contexto entendemos que a presença do termo “vandalismo” se configura como negativo e “manifestante” se apresenta como positivo e/ou neutro, temos que o portal G1 apresentou um caráter eminentemente negativo, tendo em vista que a frequência do termo vandalismo foi de 34% e manifestante foi de 11%. Por outro lado, no site de notícias Estadão foi de apenas 15% e 5%, respectivamente.

Considerações Finais

Dentre as hipóteses formuladas no início deste trabalho, temos que todas se concretizaram. No âmbito do local concluí-se que a maior parte dos enquadramentos estava focado nas manifestações ocorridas na região Sudeste, reforçando a ideia de dominação quando da produção de conteúdo a partir dos principais meios de comunicação do país, os quais estão instalados na região citada. Tal aparato contribui assim para fortalecer as coberturas feitas por meio dos veículos de referência. Durante o período analisado tivemos 3.174 notícias relacionadas aos quatro estados que compõe a região sudeste, enquanto na região nordeste tivemos apenas 691 notícias. Por outro lado, os números encontrados podem ser explicados pelo fato de que a eclosão das manifestações deu-se de maneira mais acentuada no Estado de São Paulo.

No que concerne à hipótese referente ao momento da notícia, flagrou-se que a cobertura noticiosa pautou-se, de forma mais pontual, no momento em que o evento acontecia, sendo que o portal Estadão apresentou uma cobertura mais instantânea. Das expressões mineradas, 37% delas traziam em suas notícias a palavra agora, fato este que corrobora com as ideias de imediaticidade e produção contínua cultivadas pelo meio online.

Já quanto ao caráter das notícias analisadas, detectou-se, essencialmente, um viés negativo. Fato este que pode ser confirmado pela presença constante do termo vandalismo



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





nos dois portais analisados, sendo que para o G1 o percentual apresentado foi bem maior que do portal Estadão.

Referências Bibliográficas

ARANHA, C. N. (2007). **Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional**, Tese de Doutorado, Departamento de Engenharia Elétrica, PUC- Rio.

ARANHA, C. N., & PASSOS, E. P. (2006). **A Tecnologia de Mineração de Textos**. Revista Eletrônica de Sistemas de Informação , 2, 2.

BARBOSA, S. **Jornalismo Digital em Base de Dados (JDBD) - Um paradigma para produtos jornalísticos digitais dinâmicos**. 2007. (Tese de Doutorado). PósCOM/UFBA. Disponível em: <http://www.facom.ufba.br/jol/pdf/tese_suzana_barbosa.pdf>.

BATESON, G. **Steps to an Ecology of Mind**, Northvale, New Jersey, London, Jason Aronson Inc. 1972.

CARRILHO, J. (2007). **Desenvolvimento de uma Metodologia para Mineração de Textos**. Dissertação de Mestrado, Departamento de Engenharia Elétrica, PUC-Rio.

EBECKEN, N.F.F.; LOPES, M.C.S.; COSTA, M.C.A **Mineração de Textos**. In: **Sistemas Inteligentes: Fundamentos e Aplicações**. Barueri, SP: Manole, Cap. 13, 2003.

ERBOLATO, M. **Técnicas de codificação em Jornalismo**. São Paulo: Ática, 1991.

FELDMAN, R., & SANGER, J. (2007). **The Text Mining Handbook** – Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press.

FIDALGO, António. **A resolução semântica no jornalismo online**. In: BARBOSA, Suzana (Org.). **Jornalismo digital de terceira geração**. Coleção Estudos em Comunicação. Covilhã, PT: LabcomBooks, 2007b. p. 93-102.

GOLDSCHMIDT, R., & PASSOS, E. (2005). **Data Mining: um guia prático**. Rio de Janeiro: Campus.



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com





GONÇALVES, Telmo. A abordagem do enquadramento nos estudos do Jornalismo. In: **Caleidoscópio** – Revista de Comunicação e Cultura, nº 5, 2005. Disponível em: <<http://revistas.ulusofona.pt/index.php/caleidoscopio/article/view/2268>>. Acesso em: 10 de agosto de 2013.

LIMA JR. Walter Teixeira. **Jornalismo Inteligente na era do data mining**. In: Revista do Programa de Pós-graduação da Faculdade Cásper Líbero, ano IX – no.18, p. 121-126, 2006.

NAVEGA, Sérgio. **“Princípios essenciais do data mining”**. Disponível em <http://www.intelliwise.com/snavega>. Agosto de 2002. Acessado em 05/10/2012. realizado em setembro de 2005. Disponível em: <<http://repositorio.portcom.intercom.org.br/dspace/bitstream/1904/18029/1/R0199-1.pdf>>. Acessado em: 13 de agosto de 2013.

SADÁBA, Teresa. **Framing: El encuadre de las noticias**. El binômio terrorismo-medios. 1º Ed, Buenos Aires: La Crujía, 2007.

SILVEIRINHA, Maria João. O lançamento da moeda européia e seus enquadramentos na In: XXVIII Congresso Brasileiro de Ciências da Comunicação, Rio de Janeiro, de 5 a 9 de setembro de 2005. Disponível em: <<http://www.portcom.intercom.org.br/pdfs/19406474930586936308765905671671926431.pdf>>. Acessado em: 05 de agosto de 2013.

TUCHMAN, G. **Making News**. Free Press, New York, 1978.

TAN, A. **Text Mining: The stage of the art and the challenges**. Workshop on Knowledge Discovery from Advanced Databases, Beijing, 1999

TRÄSEL, Marcelo. **Mineração de dados e textos e suas possibilidades aplicadas ao processo de produção da notícia**. Disponível em: http://sbpjour.kamotini.ghost.net/sbpjour/admjour/arquivos/ind_.pablo_barbosa.pdf Acesso: 01 outubro de 2013.

TURBAN, Efraim; R. JR, Kelly Rainer; POTTER, Richard E. **Administração de tecnologia da informação**. Rio de Janeiro: Elsevier, 2005.

WEISS. S.M . et al. **TEXT MINING**. Predictive Methods for Analyzing Unstructured Information. Springer, New York.



Grupo de Pesquisa em Ciberjornalismo – CIBERJOR-UFMS
Cidade Universitária, s/n - Caixa Postal 549 Tel: (67) 3345-7040
CEP 79070-900 * Campo Grande (MS) *
www.ciberjor.ufms.br - ciberjornalismo.ufms@gmail.com

